



Research article

Screening coronavirus and human proteins for sialic acid binding sites using a docking approach

Chia-Wen Wang¹, Oscar K. Lee^{2,3,*} and Wolfgang B. Fischer^{1,*}

¹ Institute of Biophotonics, School of Biomedical Science and Engineering, National Yang Ming Chiao Tung University, Taipei, Taiwan

² Institute of Clinical Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan

³ Department of Orthopedics, China Medical University Hospital, Taichung, Taiwan

* **Correspondence:** Email: oscarlee9203@gmail.com; wfischer@ym.edu.tw, Tel: +886228267394; Fax: +886228235460.

Abstract: The initial step of interaction of some pathogens with the host is driven by the interaction of glycoproteins of either side *via* endcaps of their glycans. These end caps consist of sialic acids or sugar molecules. Coronaviruses (CoVs), including severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), are found to use this route of interaction. The strength and spatial interactions on the single molecule level of sialic acids with either the spike (S) protein of SARS coronaviruses, or human angiotensin-converting enzyme 2 (ACE2) and furin are probed and compared to the binding modes of those sugar molecules which are present in glycans of glycoproteins. The protocol of using single molecules is seen as a simplified but effective mimic of the complex mode of interaction of the glycans. Averaged estimated binding energies from a docking approach result in preferential binding of the sialic acids to a specific binding site of the S protein of human coronavirus OC43 (HCoV-OC43). Furin is proposed to provide better binding sites for sialic acids than ACE2, albeit outweighed by sites for other sugar molecules. Absolute minimal estimated binding energies indicate weak binding affinities and are indifferent to the type of sugar molecules and the proteins. Neither the proposed best binding sites of the sialic acids nor those of the sugar molecules overlap with any of the cleavage sites at the S protein and the active sites of the human proteins.

Keywords: sialic acids; sugar molecules; spike protein; SARS-CoV-2; human receptor proteins; docking approach

1. Introduction

Interaction of sialic acids as the end caps of glycans from host glycoproteins with membrane proteins of the pathogen are considered as an initial event of the viral infectivity cycle of many viruses [1–3], e.g. human parainfluenza virus type 2 (HPF3) [4] influenza viruses [5] and also SARS coronaviruses [6]. Cleavage of the sialic acids from the glycoproteins [7], blocking sialic acids directly by carbohydrate binding agents (CBAs) [8,9], or developing competitive blockers of sialic acids binding sites on the target protein [10,11] are seen as potential routes for antiviral therapy. In this respect, identification of sialic acid binding sites and providing estimated binding energies are a key feature to support drug development.

Spike (S) glycoprotein from coronavirus is the key protein to enter the host cell, including human coronaviruses causing the common cold, as well as SARS-CoV, MERS-CoV, and also SARS-CoV-2 [12]. The protein is a homotrimeric class I fusion protein [13] built of 1,300 amino acids adopting a rod like shape of about 10 nm length. From the 22 glycosylation sites per protomer of the S protein trimer of SARs-CoV-2, about half of the sites contain fucose and 28 % mannose as end caps of the glycans, while 15 % of the sites are found to contain at least one sialic acid residue [14]. Unlike SARS-CoV S protein, S protein from SARS-CoV-2 needs to be pre-cleaved into two subunits, S1 and S2, a process which is called priming [15,16]. The priming is done in the host cell by furin protein in the ERGIC (endoplasmic-reticulum-to-Golgi intermediate compartment) [17]. SARS-CoV and HCoV-OC43 S proteins are cleaved by transmembrane serine protease 2 (TMPRSS2) at the site of the host cell plasma membrane [17]. Out of the four domains, A to D, domains A and B are involved in binding to sialic acid and the human host cell receptor angiotensin-converting enzyme 2 (ACE2), respectively, as the initial step of host cell invasion [18,19]. Upon binding to receptor, the entry process is then completed by the activity of the transmembrane S2 subunit initiating the fusion process [13]. Although SARS-CoV does not have the ability to bind sialic acid, experimental findings indicate a binding site for 9-O-Ac-Me-Sia within domain A of the HCoV-OC43 S protein [20,21]. Fast on and off rates of 9-O-acetylated sialic acid supports the idea of binding *via* avidity.

Some evidences propose SARS-CoV-2 S protein can interact with the cell-surface *via* glycans containing either heparan sulfates [22] or sialic acids [23], which is the initial stage of binding. Computational experiments suggest binding of sialic acids and derivatives to a homologous site at the domain A of SARS-CoV-2 [24].

Furin is a serine endoprotease recognizing a R-X-X-R/K sequence motif and cleaves proproteins to activate them [25]. It is a transmembrane protein with three putative glycosylation sites especially active in the secretory pathway of cells, in particular the trans-Golgi network (TGN). The enzymatic reaction is catalyzed *via* a catalytic triad of specifically oriented amino acids such as serine, histidine and aspartate [26,27]. Regarding the invasion of epithelia cells by SARS-CoV-2, activity of furin within the ERGIC of the infected cell by priming the spike protein of SARS-CoV-2 is seen as an important step to make the S-protein susceptible to interaction with the host cell receptor ACE2 [28]. Furin is seen as a potential target for drug development due to its essential role in the life cycle of the virus [25].

The zinc metallopeptidase ACE2 is a receptor glycoprotein specially expressed in epithelia cells and involved in regulating hypertension [29]. Especially in lungs epithelia cells it allows for an entry

passage for the virus [30]. Dimeric ACE2 has a high affinity binding site for the S protein of SARS-CoV-2 which makes it the prime reason for the infectivity of the virus [28]. Binding of the S protein to ACE2 readies TMPRSS2 to S protein priming [31]. ACE2 hosts seven glycosylation sites per protomer which also contain sialic acids [32]. Experiments in which sialic acids are removed from the glycans show rather an enhanced binding affinity to the S protein of SARS-CoV-2 than a weakening of the binding indicating a minor role of the sialic acids upon binding with the S protein.

In this study spatial distribution and estimated binding energies of sialic acids to either S protein, ACE2 or furin as negatively charged monomers are investigated. The data are compared to the binding properties of other sugar molecules present in glycans. Available crystal structures of the three proteins are taken as targets. The results are discussed in terms of their implication of the viral infectivity cycle. The questions addressed are (i) whether one of the proteins has a preferred binding site for sialic acids over the other proteins and (ii) whether the estimated binding energies of sialic acids are preferential over the energies of the sugar molecules.

Single sugar molecules are taken as a model system for probing the mode of interaction of the sialic acids with the proteins. Visible inspection is used to screen the poses of the sugar molecules for proper orientation so that they could be linked to the glycan chain of the protein *via* a α 2-6 linkage. In such a pose the linking hydroxy moiety is facing the environment rather than being oriented towards the protein. The interactions of the sialic acid are compared with those of the other sugar molecules which exist in the glycans and are identified as potential end caps.

2. Experimental section

The protein structures were taken from the Protein Data Bank (www.rcsb.org): domains A of human coronavirus OC43 (HCoV-OC43) spike (S)-proteins (PDB ID: 6NZK) including ligand (l) 9-*O*-AC-Me-Sia (Sp-l), HCoV-OC43 (PDB ID: 6OHW) with no ligand (Sp), S protein of SARS-CoV-2 (PDB ID: 6VSB) as the original S-protein of CoV-2, (Sp^{2or}), as well as the peptidase domain of human proteins angiotensin-converting enzyme 2 (ACE2) (PDB ID: 6LZG), and serine-protease furin (PDB ID: 6HZA), which includes in its holo form the synthetic peptide RRKR-Amba (furin-h). Furin-h has also been used in its apo form (furin-a)

The protein structures were either used (i) as experimental structures from the protein data bank without any minimization protocol (mp-0), (ii) structures with the side chains minimized but backbone atoms restraint (mp-1), or (iii) fully minimized structures with both, side chain and backbone atoms minimized (mp-2). Short minimization (using Molecular Operation Environment (MOE) suit, www.chemcomp.com) was done by applying steepest descent, as well as consequent conjugated gradient and truncated-Newton calculations using the Amber10 force field.

The structures of the sugar molecules N-acetylgalactosamine (GalNAc), N-acetylglucosamine (GlcNAc), the sialic acids N-acetylneuraminic acid (Neu5Ac), and N-Glycolylneuraminic acid (Neu5Gc) as well as the sugar molecules fucose, galactose, mannose, were taken from PubChem (pubchem.ncbi.nlm.nih.gov) and generated using the MOE suit. The sialic acid methyl 9-*O*-acetyl-sialic acid (9-*O*-Ac-Me-Sia) was obtained from PDB data bank (PDB ID: 6NZK). All structures were operated through a short minimization as described above.

2.1. Structure preparation and minimization

For docking to HCoV-OC43 (Sp-1), which included the sialic acid 9-*O*-AC-Me-Sia bound to it, the sialic acid was removed and only the sialic acid binding site was considered for docking. The amino acids of the protein forming the binding pocket were selected by applying a radius of 7 Å around each of the atom of the ligand. The same protein HCoV-OC43 but without ligand (Sp) was used in LeadIT to identify the pockets (number of pockets identified by LeadIT: 9 (mp-0), 10 (mp-1), 11 (mp-2)). Swiss-Model (<https://swissmodel.expasy.org>) was used to fill the existing gaps in Sp2^{or} *via* homology modelling with itself. This protocol generated a S-protein monomer, hither forth referred to as Sp2^h. In addition, the three subunits A, B, C of Sp2^{or} are used separately for docking. Only domain A (N-terminal domain) of each S protein was used for docking.

The furin structure contained a cyclic synthetic peptide RRKR-Amba. Docking was done with furin in its holo form (furin-h), including the RRKR peptide from which the Amba appendix was deleted, as well as in the absence (furin-a) of this peptide. Number of pockets identified by LeadIT were 9 for both, furin-h/furin-a (mp-0), 9 (furin-h) and 12 (furin-a) (mp-1), 9 (furin-h) and 7 (furin-a) (mp-2). Both catalytic domain and P domain were used for docking.

The ACE2 structure used was part of a complex of ACE2 being in contact with the receptor-binding domain of the SARS-CoV2. The S protein was removed and ACE2 P domain was used for docking. Number of pockets identified by LeadIT were 12 using mp-0 and mp-1 structures, and 14 using mp-2 structures.

2.2. Docking software

LeadIT (BioSolveIT, Germany): Putative pockets (P) for ligand binding were suggested by the software and used for docking. The docking software was used in its default mode and the score values representing estimated free energy ΔG (kJ/mol) were taken for data analysis. With the HYDE routine, the respective values were corrected by a dehydration enthalpy.

Decoy molecules were obtained using DUD-E in combination with the ZINK database [33]. Seven independent runs were conducted from which each run generated 50 molecules. Overlapped ligands found in several runs were deleted and finally 300 ligands ranked by the ZINK database entry number were chosen.

2.3. Analysis software

Receiver operating characteristic curve (ROC) and area-under the curve (ARC) calculations are done using Origin9 (OriginLab Corporation, Northampton, MA).

3. Results

The docking approach follows the intention to mimic interaction of sialic acids of the S proteins with the human receptor proteins ACE2 and furin (Figure 1). The estimated binding energies of the S proteins correspond to those sialic acids which are located at the sites of the human proteins, while those values of the human proteins correspond to those values of the sialic acids at the sites of the S proteins. The same scheme applies to the estimated binding energies of the sugar molecules.

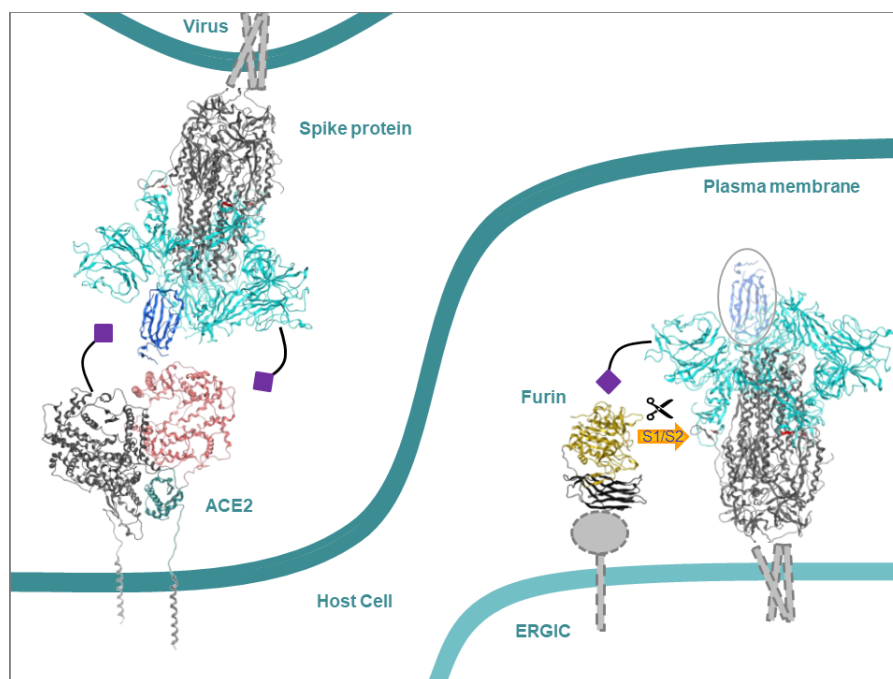


Figure 1. Schematic representation and spatial orientation of the SARS-CoV-2 S (PDB ID: 6VSB) protein and the human proteins ACE2 (PDB ID: 6M17) and furin (PDB ID: 6HZA). The cyan and grey parts of the S protein show the S1 and S2 domains, respectively. The A domain is shown in blue, while the other two domains are not shown for clarity. The encircled A domain of the S protein should not stand up in the ERGIC and is therefore shaded in grey. The two domains in ACE2 are shown in red (peptidase domain) and olive green (collectrin-like domain) while the other protein of the dimer is shown in grey. For furin, its two domains are shown in yellow (catalytic domain) and black (P domain). Undetermined structures of all the proteins are depicted in grey dashed shapes. The glycans of the S protein and ACE2 are marked as lines with purple diamonds. The S1/S2 cleavage site (scissors) of the S protein is marked with an arrow.

3.1. Binding affinities from the average over various binding sites

A total of eight sugar molecules including 3 sialic acids and 5 sugar molecules which are commonly used as the building block of polysaccharides, are docked to various S protein structures as well as to ACE2 and furin (Figure 2). The values for each of the sugar molecules are a result of averaging over (i) the structures derived from the three minimization protocols (mp-0, mp-1, mp-2), (ii) the best binding energies in the individual pockets which are identified by the docking software of the respective protein, and (iii) the three sialic acids (Figure 3, Sia) and accordingly over the five sugar molecules (Figure 3, Su). Docking results are shown for the experimentally derived structures for Sp-1 without the ligand 9-*O*-AC-Me-Sia at the binding site. In addition, the afore mentioned results are averaged further over the three protein structures Sp, Sp2^h and Sp2^{or}, as well as the two structures furin-a and furin-h.

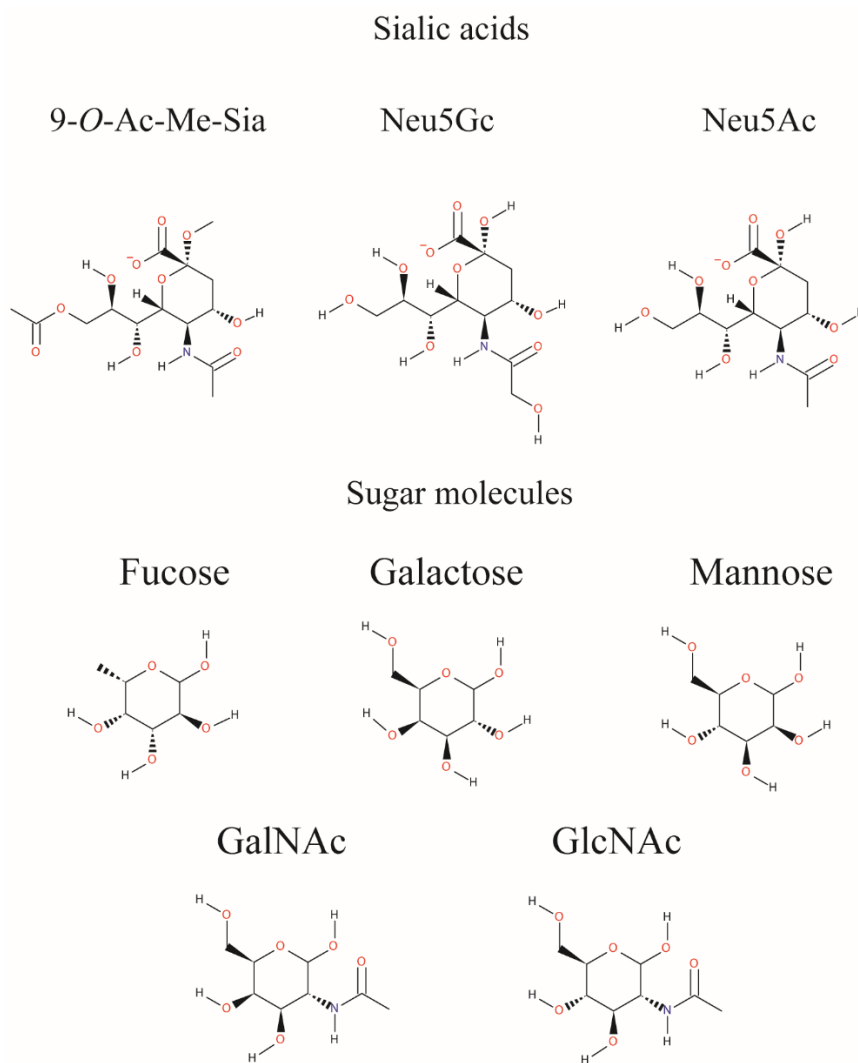


Figure 2. Chemical structures of the sialic acids and sugar molecules in glycans of glycoproteins. (Oxygen atoms are shown in red.)

The top 20 poses are inspected and used for data analysis. Data analysis are based on either the sialic acids and the sugar molecules ranked as number one (rank-1) or the best scored and oriented sugar molecules (oriented). The oriented sugar molecule is identified by visible inspection opting for an orientation of the respective O-sites on the sugar molecules so that the O-sites could be sterically linked to a putative glycol-chain. In case that none of the 20 poses show an adequate orientation for being selected as ‘oriented’, the finding was considered as having no-docking pose.

The binding energies are the lowest (sialic acids (-20.1 ± 1.6) kJ/mol, sugar molecules (-16.7 ± 0.7) kJ/mol for the oriented ligands docked to Sp-1 compared to those values obtained to docking them to the other experimental structures (Figure 3A, blue bars, and Suppl. Table 1a). The sites at the furin-a/h reveal values as low as (-11.0 ± 0.5) kJ/mol for sialic acids and (-13.7 ± 0.3) kJ/mol for the sugar molecules. Correcting the poses for hydration penalty using the HYDE routine does not change the pattern between sialic acids and sugar molecules (Figure 3, orange bars and Suppl. Table 1a).

Averaging over the rank-1 ligands, reveals a similar pattern as mentioned for the oriented sugar molecules with values calculated for Sp-1 as (-20.6 ± 1.4) kJ/mol for sialic acids and (-17.3 ± 1.3) kJ/mol for the sugar molecules (Figure 3B, and Suppl. Table 1b). The respective values for furin unfold as (-13.3 ± 0.7) kJ/mol for sialic acids and (-14.4 ± 0.3) kJ/mol for the sugar molecules.

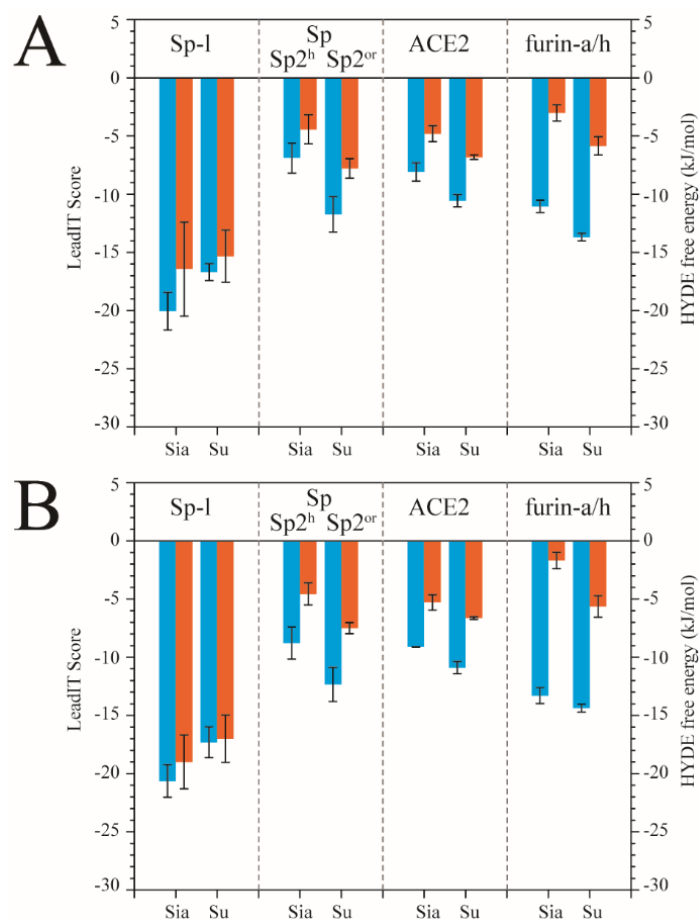


Figure 3. Averaged estimated binding energies of sialic acids and sugar molecules to S protein structures as well as human proteins ACE2 and furin. (A) averaged values from oriented best scored poses and (B) those for the best ranked molecules. Blue bars show the values for LeadIT, the orange bars for the respective best position according to HYDE values. Sia and Su stand for sialic acids and sugar molecules, respectively. Standard deviations (SD) for each, Sp-1 and ACE2, are taken from values averaged over same ligand types (sialic acids (Sia) or sugar molecules (Su)) in the same pocket identified by the software, consequently averaging over the number of pockets for each type is done, followed by an average over the three different protocols mp-0 to mp-2. In the case of the S proteins and furin there is another average over the different protein types, e.g. Sp, Sp2^h and Sp2^{or} for the S proteins and apo (a) and holo (h) form for furin.

The sialic acids bind most strongly to the experimentally identified binding site followed by strong binding sites at furin proteins. When using available S-protein structures without the identified specific site estimated binding energies are indifferent with those values derived for the human

proteins. There are only a few structural features for which the sialic acids bind better than the sugar molecules. In many cases the sugar molecules are at least equally, good binders as the sialic acids.

3.2. The individual docking sites

3.2.1. Maximum difference values between sialic acids and sugar molecules

The respective difference between the estimated binding energies of sialic acids and sugar molecules are an average taken over the largest difference identified for each pocket of the protein and consequent averaging over either the 3 sialic acids or the 5 sugar molecules. (Suppl. Table 2)

Identifying the biggest difference values on the basis of the LeadIT values for each of the proteins shows that the values for the oriented molecules are as high as (6.1 ± 0.8) kJ/mol for (Sp-l), (5.2 ± 2.1) kJ/mol for ACE2 and (6.9 ± 1.5) kJ/mol for furin-a (Suppl. Table 2). Selecting the rank-1 poses, the numbers are almost similar with (6.0 ± 1.2) kJ/mol for Sp-l, (6.2 ± 1.5) kJ/mol ACE2, and (6.5 ± 1.5) kJ/mol for furin-a.

Focusing on the HYDE values, the differences of the values for the oriented poses are as high as (14.5 ± 4.1) kJ/mol for Sp^{2or}, (8.9 ± 3.0) kJ/mol for ACE2, and 13.5 kJ/mol for both furin proteins furin-h (standard deviation (SDEV) ± 5.5 kJ/mol) and furin-a (STDEV ± 6.7 kJ/mol) (Suppl. Table 2). The respective HYDE values for the rank-1 poses are (15.7 ± 4.8) kJ/mol for Sp^{2or}, (12.8 ± 6.2) kJ/mol for ACE2 and (13.9 ± 6.7) kJ/mol for furin-a.

The difference between sialic acid and sugar binding energies can be as high as 6–7 kJ/mol (LeadIT) and 14–15 kJ/mol (HYDE). There is also a trend that the differences for the human proteins are slightly higher than the differences for the viral proteins.

3.2.2. Best binding sites

All absolute values for both, the oriented and rank-1 poses, are in the range of -20 to -30 kJ/mol (Table 1 and Suppl. Table 3). Searching for the absolute lowest oriented values obtained from all docking poses for the individual proteins reveals that the values do not differ between the sialic acids and sugar molecules ((-23.9 ± 0.7) kJ/mol (sialic acids) *versus* (-23.9 ± 3.4) kJ/mol (sugar molecules)) for the S-protein structures and those of the human proteins ((-23.5 ± 2.2) kJ/mol (sialic acids) *versus* (-21.3 ± 0.9) kJ/mol (sugar molecules)) for LeadIT (Table 1). For the S-proteins in two cases, Sp and Sp^{2h}, the sugar molecules have lower values than the sialic acids. In the case of the human proteins, it is Neu5Ac which adopts the lowest value (e.g. Neu5Gc to ACE2: -24.9 kJ/mol). There is no difference between the energies of the sialic acids at the S proteins and the human proteins.

The pattern is similar if the rank-1 poses are selected (Suppl. Figure 2). The difference between sialic acids and sugar molecules of the S-protein is marginal with values of (-24.1 ± 0.5) kJ/mol (sialic acids) *versus* (-24.6 ± 2.8) kJ/mol (sugar molecules), but slightly enhanced when comparing the energies amongst the human proteins ((-28.2 ± 2.1) kJ/mol (sialic acids) *versus* (-23.1 ± 1.4) kJ/mol (sugar molecules)). Amongst the sialic acids, the estimated binding energies are lower for binding to the human proteins than for the sugar molecules.

Table 1. Oriented sugar molecules with the lowest estimated binding energies identified for LeadIT and HYDE. ‘+’: sialic acid binding poses which are the same for both oriented and rank-1; ‘+++’: best docked sialic acids and used for decoy finding. Red: lowest value observed over all protocols and pockets.

	LeadIT	Sialic acids	Sugar molecules		
			ΔG (kJ/mol)	ΔG (kJ/mol)	
S-proteins	Sp-1	+/ +++Neu5Gc	24.5	Gal	20.7
	Sp	Neu5Gc	24.4	Gal	26.9
	Sp2 ^h	Neu5Gc	23.2	GlcNAc	26.8
	Sp2 ^{or}	Neu5Ac	23.3	Fucose	21.2
	avg. ΔG		23.9 ± 0.7		23.9 ± 3.4
Human proteins	ACE2	++Neu5Gc	24.9	Gal	21.8
	furin-h	Neu5Gc	20.9	GalNAc	20.2
	furin-a	++Neu5Gc	24.6	GalNAc	21.8
	avg. ΔG		23.5 ± 2.2		21.3 ± 0.9
<hr/>					
	HYDE	Sialic acids	Sugar Molecules		
			ΔG (kJ/mol)	ΔG (kJ/mol)	
S-proteins	Sp-1	Neu5Ac	25.0	Mannose	25.0
	Sp	Neu5Gc	28.0	Galactose	28.0
	Sp2 ^h	Neu5Ac	25.0	GlcNAc	33.0
	Sp2 ^{or}	+/ +++Neu5Gc	31.0	GlcNAc	27.0
	avg. ΔG		27.3 ± 2.9		28.3 ± 3.4
Human proteins	ACE2	++Neu5Ac	26.0	Mannose	25.0
		+/ +++9-O-AC-Me-Sia	28.0	Mannose	24.0
	furin-h	Neu5Gc	28.0		
	furin-a	9-O-AC-Me-Sia	26.0	GalNAc	30.0
	avg. ΔG		27.0 ± 1.2		26.3 ± 3.2

For the S proteins Neu5Gc (−31.0 kJ/mol) scores highest for the oriented poses looking at the HYDE values (Table 1). When average over the best poses of all the S protein the binding energies are slightly in preference for the sugar molecules ((−27.3 ± 2.9) kJ/mol (sialic acids) *versus* (−28.3 ± 3.4) kJ/mol (sugar molecules) ($p = 0.7$)). The lowest values are obtained for the sialic acids 9-O-AC-Me-Sia and Neu5Gc on furin-h (−28.0 kJ/mol) followed by Neu5AC -26.0 kJ/mol at ACE2. The HYDE values of the sialic acids binding to the human proteins are not better than those values for binding to the S protein ($p = 0.9$). The overall pattern as described remains the same when selecting the rank-1 ligands, except that the lowest estimated binding energy for Neu5Gc (−31.0 kJ/mol) with ACE2 is followed by energy values of Neu5Gc and 9-O-AC-Me-Sia (−28.0 kJ/mol) with the furin proteins (Suppl. Table 3).

In terms of absolute estimated binding energies, the binding energies do not differ so much

between the sialic acids and the sugar molecules. When collecting the data over a series of the different protein conformations, the average values are in preference for the sialic acids.

3.3. Binding poses

The positions of the best poses of the sialic acids on the S-protein coincide with the site identified in the structure of Sp¹ (Figure 4). In case of the human proteins, some sites are close to the identified active sites such as the catalytic site of furin (e.g. orange site in Figure 4B and site in Figure 5B, III)).

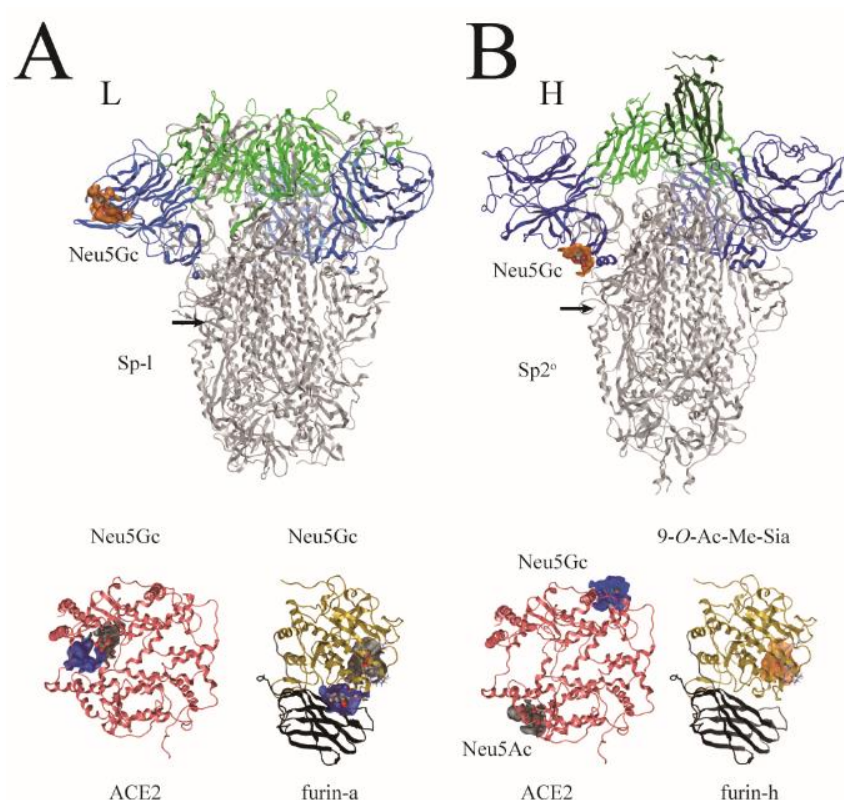


Figure 4. Location of the poses with the best estimated binding energies on the surface of the individual proteins. (A) Poses are shown based on the values derived from LeadIT (L) and (B) the respective HYDE (H) values. The orange-colored pockets: pocket and binding pose for the oriented and rank-1 sialic acids are the same; blue pockets: oriented sialic acids; grey pockets: rank-1 sialic acids. The black arrow indicates the S1/S2 cleavage site.

The poses of the best oriented sialic acids are stabilized by 8–9 hydrogen (h) bonds (9 h-bonds for Sp-1 and furin-a; 8 h-bonds for ACE2) (Figure 5A) and a slightly lower number of hydrogen bonds for the best poses identified by HYDE (6 h-bond for all proteins) (Figure 5B). Additional other hydrophilic residues such as serine and threonine as well as asparagine and glutamine residues are in close contact with the sialic acids. This pattern is also found for the rank-1 positions (10 h-bonds for Sp-1 and furin-a, 11 h-bonds for ACE2 for LeadIT (Suppl. Figure 1A) and 6 h-bonds for Sp2^{or} and

furin-h as well as 8 h-bonds for ACE2) for HYDE (Suppl. Figure 1B). HYDE poses show lower number of h-bonds than those for LeadIT which is in concert with the idea of considering dehydration upon binding as a penalty. In many cases the best sites according to the docking results of LeadIT, Neu5Gc Sp-1, and HYDE, Neu5Gc Sp2^{or}, 9-O-Ac-Me-Sia furin-h, are identical for the oriented and the rank-1 ligands.

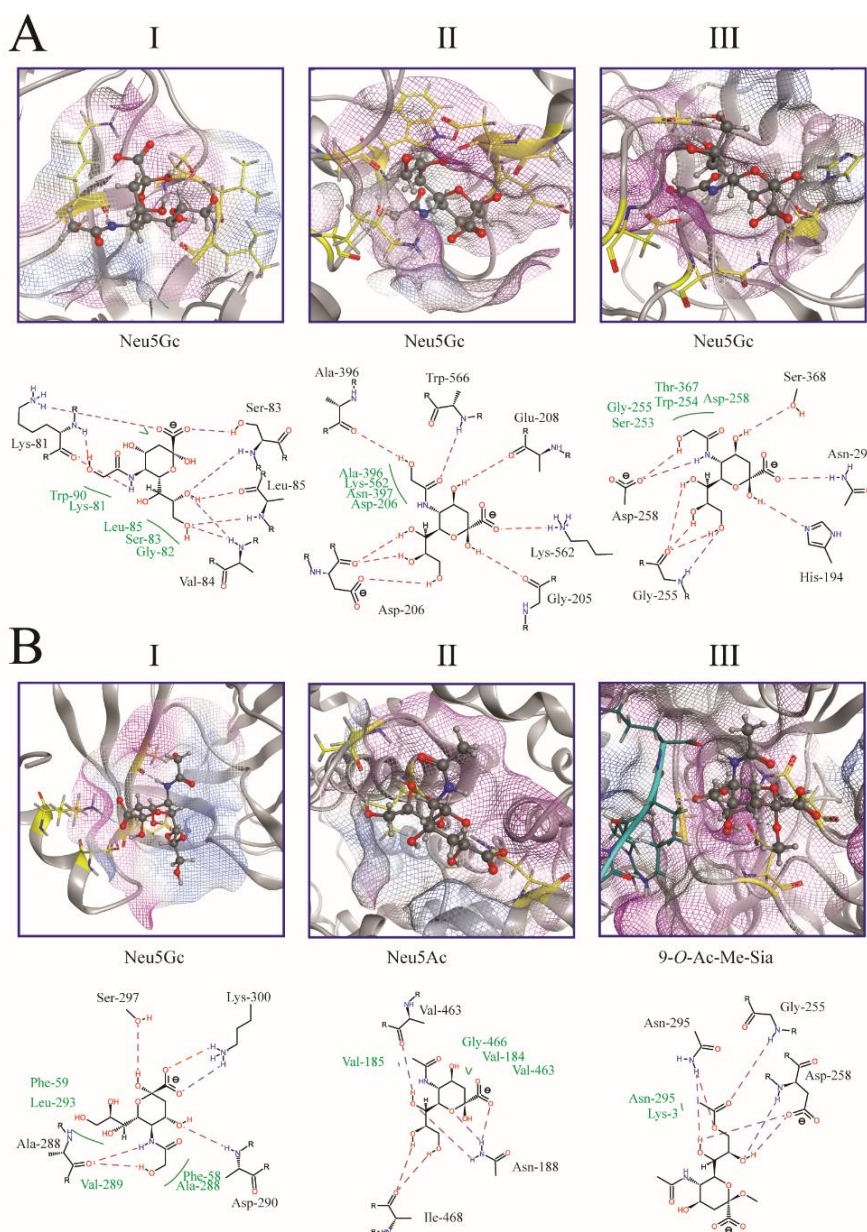


Figure 5. Absolute best binding poses of oriented sialic acids and sugar molecules. (A) from left to right based on the LeadIT scoring the binding sites of Neu5Gc to protein structure (I) Sp-1, (II) ACE2 and (III) furin-a. (B) in the similar sequence based on their scoring using the best HYDE values from left to write the binding site of (I) Neu5Gc to Sp-1, (II) Neu5Ac to ACE2 and (III) 9-O-Ac-Me-Sia to furin-h. The structural features in the sugar in the binding pocket are shown in the upper row and the respective 2D maps of the pocket in the lower row. The van der Waals mesh colored in pink represents

hydrophilic surfaces, the blue mesh represents hydrophobic surfaces. The light blue peptide indicates the catalytic site of furin-h. See Table 1 for the respective estimated binding energies.

3.4. Specificity of binding sites

The best binding sites in terms of absolute binding energies for all three proteins are evaluated using 300 decoys. All decoys chosen for the individual sugar molecules show the highest contribution of molecules with 5–7 rotatable bonds (Suppl. Figure 3). Due to 9-*O*-AC-Me-Sia (8 rotatable bonds) some decoys also have up to 11 rotatable bonds. Thus, they are chosen in the range of rotatable bonds identified for the sugar molecules which is not exceeding the number of 8 rotatable bonds (Suppl. Table 5).

The AUC values for the oriented poses for LeadIT sample from the ROC plot follows Sp-1 > furin-a > ACE2 (rank-1: ACE2 > Sp-1 > furin-a (Suppl. Figure 2A and Suppl. Table 5). Looking at the best respective HYDE values the sequence remains the same, however the AUC numbers are high than for the LeadIT values (oriented Sp-1 > furin-a > ACE2 (rank-1: ACE2 > Sp-1 > furin-a) (Suppl. Table 6).

The AUC values for the oriented and ranked-1 poses for the absolute best HYDE samples are above 0.99 independent of the protein showing a fully specific binding of the sugar molecules (Suppl. Figure 2B and Suppl. Table 5). The AUC values for the corresponding LeadIT values are all below 0.1 with those for furin-h the highest of 0.07 for both oriented and ranked-1 poses compared to 0.04 (oriented) and 0.06 (rank-1) for ACE2 and 0.02 (oriented/rank-1) for Sp2^{or}. (Suppl. Table 6)

Whilst scoring by LeadIT suggests that the sites are less specific for the individual sugar molecules, the poses scored by HYDE suggest highly specific sites.

4. Discussion

In this study the LeadIT is used to perform docking of the sialic acids and sugar molecules. The software fragmentizes the molecule at rotatable bonds and reassembles the molecule within the identified binding pockets. It shows good performance [34] and has been used in docking approaches of dipeptidyl peptidase-IV (DPPIV) inhibitors rich in hydroxy groups [35]. The number of rotatable bonds of the sugar molecules is in the range of 0 (fucose) to 8 (9-*O*-AC-Me-Sia (Suppl. Table 4) and thus, below a reported threshold of ≥ 10 which is reported to lower the performance of the software [36,37]. No preferred binding topology are considered letting the program placing the sugar molecules without constrains like reported for docking of sugar molecules to e.g. Ca-dependent lectins [38].

4.1. Interpretation of the results and biological interpretation

A putative binding site of a sialic acid derivative to the S protein is identified experimentally and supported by the docking data. The presented data propose, that this site is a better sialic acid binding site than a sugar binding site. No other site on the A domain of S protein seems to be able to compete with this site in terms of absolute values or whether sialic acids are preferred over sugar molecules. SARS-CoV-2 binds to human proteins, here e.g. ACE2, *via* their sialic acids is in preference over binding of their sugar molecules. Also, the best estimated binding energy of the

sialic acids on a human protein is found for furin, making the interaction of SARS-CoV2 S protein with furin in the ERGIC [17] a moderately enhanced interaction.

Absolute binding numbers including those for HYDE are within a narrow range (28–34 kJ/mol) and with these values, corresponding to an apparent binding constant in the micro molar range, indicating that the sugar molecules are moderate binders [39]. This finding is in accordance with results from experiments in which sialic acids, when connected with the sugar molecules and interact with e.g. CD45 [40] or with sialic acid-binding immunoglobulin-type lectins (siglecs) [41,42], are classified as weak binders with binding affinities in the range of 0.1–3 mM.

While the LeadIT sites are not very specific for the individual sialic acid, the HYDE sites are very specific. Since the differences between the binding energies of the sialic acids and sugar molecules are marginal, this finding supports the idea that the accumulation of the individual binding sites for the polysaccharide chains on the target protein with its sialic-acid caps drives binding *via* avidity rather than affinity.

Binding of negatively charged sialic acids of host glycoproteins to the S proteins is the entry route of some coronaviruses besides the proteinaceous interaction with the receptors [43]. The S protein can be the target of the polysaccharides of the human proteins, in as much human ACE2 has 7 potential N-glycosylation sites and 3 O-glycosylation sites [44], furin harbours only three potential N-glycosylation sites [45]. On the other hand, the viral S protein of HCoV-OC43 has 23 glycosylation sites (22 for SARS-CoV-2 [46]) sites with 12 sites confirmed to be glycosylated [47,48] and about 15 % of these sites containing sialic acids [14]. The modest good numbers for sialic acids binding to furin proposes that SARS-CoV-2 S protein will target the furins more thoroughly than the ACE2s which is proposed to increase the efficiency of furin cleavage. It is proposed that this binding of the S protein to the furin is as efficient as binding to ACE2s [17].

Sialic acid binding sites may not interfere with the active sites but are close to active sites, supporting the idea of their role as an anchoring tool locking the approaching proteins into the proper position for interaction.

5. Conclusions

Sialic acids have a slightly preferred binding mode over the sugar molecules. An overall weak binding affinity in combination with the number of glycosylation sites on either of the proteins, proposes that binding is driven by a combination of avidity and affinity. The weighting of whether either of the scenario prevails could depend on the strength of the affinity or putative number of glycosylation sites on the proteins.

Binding of the SARS-CoV-2 S protein to furin is proposed to be due to affinity rather than avidity, based on somewhat higher binding affinities in combination with the lower number of glycosylation sites on the protein. This scenario is suggested to be reversed in the case of binding of the protein to ACE2 where the moderate strength of the interaction is compensated by avidity.

Acknowledgments

WBF and OKL thank the Ministry of Science and Technology, Taiwan (MOST-107-2112-M-010-001-MY3 to WBF and MOST 109-2823-8-010-003-CV to OKL) for financial support. We are grateful to the National Center for High-performance Computing, Hsinchu, Taiwan, for

computer time and facilities.

Conflicts of interest

The authors declare there is no conflict of interest.

References

1. Varki A (2008) Sialic acids in human health and disease. *Trends Mol Med* 14: 351–360.
2. Stencel-Baerenwald JE, Reiss K, Reiter DM, et al. (2014) The sweet spot: defining virus-sialic acid interactions. *Nat Rev Microbiol* 12: 739–749.
3. Koehler M, Delguste M, Sieben C, et al. (2020) Initial step of virus entry: virion binding to cell-surface glycans. *Annu Rev Virol* 7: 143–165.
4. Moscona A, Peluso RW (1993) Relative affinity of the human parainfluenza virus type 3 hemagglutinin-neuraminidase for sialic acid correlates with virus-induced fusion activity. *J Virol* 67: 6463–6468.
5. Viswanathan K, Chandrasekaran A, Srinivasan A, et al. (2010) Glycans as receptors for influenza pathogenesis. *Glycoconj J* 27: 561–570.
6. Sun XL (2021) The role of cell surface sialic acids for SARS-CoV-2 infection. *Glycobiology* online ahead of print.
7. Nicholis JM, Moss RB, Haslam SM (2013) The use of sialidase therapy for respiratory viral infections. *Antivir Res* 98: 401–409.
8. Balzarini J (2007) Targeting the glycans of glycoproteins: a novel paradigm for antiviral therapy. *Nat Rev Microbiol* 5: 583–597.
9. Colpitts CC, Schang LM (2014) A small molecule inhibits virion attachment to heparan sulfate- or sialic acid-containing glycans. *J Virol* 88: 7806–7817.
10. Rustmeier NH, Strebl M, Stehle T (2019) The symmetry of viral sialic acid binding sites—implications for antiviral strategies. *Viruses* 11: 947.
11. Heida R, Bhide YC, Gasbarri M, et al. (2020) Advances in the development of entry inhibitors for sialic-acid-targeting viruses. *Drug Discov Today* 26: 122–137.
12. Li F (2016) Structure, function, and evolution of coronavirus spike proteins. *Annu Rev Virol* 3: 237–261.
13. Bosch BJ, Van der Zee R, De Haan CAM, et al. (2003) The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J Virol* 77: 8801–8811.
14. Watanabe Y, Allen JD, Wrapp D, et al. (2020) Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* 369: 330–333.
15. Millet JK, Whittaker GR (2014) Host cell entry of middle east respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. *P Natl Acad Sci USA* 111: 15214–15219.

16. Coutard B, Valle C, de Lamballerie X, et al. (2020) The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res* 176: 104742.
17. Tang T, Bidon M, Jaimes JA, et al. (2020) Coronavirus membrane fusion mechanism offers a potential target for antiviral development. *Antivir Res* 178: 104792.
18. Li F, Li W, Farzan M, et al. (2005) Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* 309: 1864–1868.
19. Mathewson AC, Bishop A, Yao Y, et al. (2008) Interaction of severe acute respiratory syndrome-coronavirus and NL63 coronavirus spike proteins with angiotensin converting enzyme-2. *J Gen Virol* 89: 2741–2745.
20. Tortorici MA, Walls AC, Lang Y, et al. (2019) Structural basis for human coronavirus attachment to sialic acid receptors. *Nat Struct Mol Biol* 26: 481–489.
21. Hulswit RJG, Lang Y, Bakkers MJG, et al. (2019) Human coronaviruses OC43 and HKU1 bind to 9-O-acetylated sialic acids via a conserved receptor-binding site in spike protein domain A. *P Natl Acad Sci USA* 116: 2681–2690.
22. Hao W, Ma B, Li Z, et al. (2021) Binding of the SARS-CoV-2 spike protein to glycans. *Sci Bull* 66: 1205–1214.
23. Baker AN, Richards SJ, Guy CS, et al. (2020) The SARS-COV-2 spike protein binds sialic acids and enables rapid detection in a lateral flow point of care diagnostic device. *ACS Cent Sci* 6: 2046–2052.
24. Awasthi M, Gulati S, Sarkar DP, et al. (2020) The sialoside-binding pocket of SARS-CoV-2 spike glycoprotein structurally resembles MERS-CoV. *Viruses* 12: 909.
25. Wu C, Zheng M, Yang Y, et al. (2020) Furin, a potential therapeutic target for COVID-19. *iScience* 23: 101642.
26. Thomas G (2002) Furin at the cutting edge: from protein traffic to embryogenesis and disease. *Nat Rev Mol Cell Biol* 3: 753–766.
27. Anderson ED, Molloy SS, Jean F, et al. (2002) The ordered and compartment-specific autoproteolytic removal of the furin intramolecular chaperone is required for enzyme activation. *J Biol Chem* 277: 12879–12890.
28. Shang J, Wan Y, Luo C, et al. (2020) Cell entry mechanisms of SARS-CoV-2. *P Natl Acad Sci USA* 117: 11727–11734.
29. Turner AJ, Hiscox JA, Hooper NM (2004) ACE2: From vasopeptidase to SARS virus receptor. *Trends Pharmacol Sci* 25: 291–294.
30. Li Y, Zhou W, Yang L, et al. (2020) Physiological and pathological regulation of ACE2, the SARS-CoV-2 receptor. *Pharmacol Res* 157: 104833.
31. Hoffmann M, Kleine-Weber H, Schroeder S, et al. (2020) SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181: 271–280.
32. Allen JD, Watanabe Y, Chawla H, et al. (2021) Subtle influence of ACE2 glycan processing on SARS-CoV-2 recognition. *J Mol Biol* 433: 166762.
33. Mysinger MM, Carchia M, Irwin JJ, et al. (2012) Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J Med Chem* 55: 6582–6594.
34. Bursulaya BD, Totrov M, Abagyan R, et al. (2003) Comparative study of several algorithms for flexible ligand docking. *J Comput Aid Mol Des* 17: 755–763.

35. Rozano L, Abdullah Zawawi MR, Ahmad MA, et al. (2017) Computational analysis of Gynura bicolor bioactive compounds as dipeptidyl peptidase-IV inhibitor. *Adv Bioinform* 5124165.
36. Kellenberger E, Rodrigo J, Muller P, et al. (2004) Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins* 57: 225–242.
37. Chen H, Lyne PD, Giordanetto F, et al. (2006) On evaluating molecular-docking methods for pose prediction and enrichment factors. *J Chem Inf Model* 46: 401–415.
38. Nurisso A, Kozmon S, Imberty A (2008) Comparison of docking methods for carbohydrate binding in calcium-dependent lectins and prediction of the carbohydrate binding mode to sea cucumber lectin CEL-III. *Mol Simulat* 34: 469–479.
39. Reulecke I, Lange G, Albrecht J, et al. (2008) Towards an integrated description of hydrogen bonding and dehydration: decreasing false positives in virtual screening with the HYDE scoring function. *Chem Med Chem* 3: 885–897.
40. Bakker TR, Piperi C, Davies EA, et al. (2002) Comparison of CD22 binding to native CD45 and synthetic oligosaccharide. *Eur J Immunol* 32: 1924–1932.
41. Crocker PR, Blixt O, Collins BE, et al. (2003) Sialoside specificity of the siglec family assessed using novel multivalent probes: identification of potent inhibitors of myelin-associated glycoprotein. *J BiolChem* 278: 31007–31019.
42. Yamakawa N, Yasuda Y, Yoshimura A, et al. (2020) Discovery of a new sialic acid binding region that regulates Siglec-7. *Sci Rep* 10: 1–14.
43. Li W, Hulswit RJG, Widjaja I, et al. (2017) Identification of sialic acid-binding function for the Middle East respiratory syndrome coronavirus spike glycoprotein. *Proc Natl Acad Sci USA* 114: E8508–E8517.
44. Shajahan A, Archer-Hartmann S, Supekar NT, et al. (2020) Comprehensive characterization of N- and O-glycosylation of SARS-CoV-2 human receptor angiotensin converting enzyme 2. *Glycobiology* 31: 410–424.
45. Mamedov T, Musayeva I, Acsora R, et al. (2019) Engineering, and production of functionally active human Furin in *N. benthamiana* plant: *In vivo* post-translational processing of target proteins by Furin in plants. *PLoS One* 14: e0213438.
46. Chen Y, Guo Y, Pan Y, et al. (2020) Structure analysis of the receptor binding of 2019-nCoV. *Biophys Res Comm* 525: 135–140.
47. Krokhin O, Li Y, Andonov A, et al. (2003) Mass spectrometric characterization of proteins from the SARS virus: a preliminary report. *Mol Cell Proteomics* 2: 346–356.
48. Fung T S, Liu D X (2018) Post-translational modifications of coronavirus proteins: roles and function. *Future Virol* 13: 405–430.



AIMS Press

© 2021 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)